Infering the selective pressures acting on insertions and deletions in the great tit genome

Henry Barton

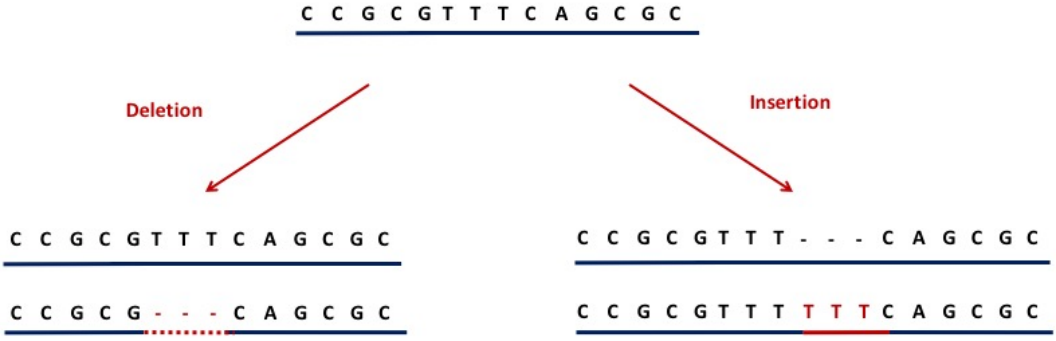
The University of Sheffield

22/08/18

Introduction

Insertions and deletions

- short INDELs: sections of DNA $<$ 50bp that are deleted or inserted in a genome
- deletion bias in most organisms
- deletions more deleterious than insertions



INDELs often overlooked

- Disproportionately occur in repetitive sequence
- Hard to align
- Often occur in hotspots
- 1/8 as frequent as SNPs in humans

(Earl et al., 2014; Montgomery et al., 2013)

The importance of INDELs in genome evolution

- Influence genome size:
  - low deletion rate $\rightarrow$ large genomes?
  - high deletion rate $\rightarrow$ compact genomes?
  - selection on insertions to maintain minimum intron size?
- Contribute more to sequence divergence, in terms of the number of base differences, than SNPs

(Britten, 2002; Leushkin and Bazykin, 2013; Nam and Ellegren, 2012; Ometto et al., 2005; Sun et al., 2012)

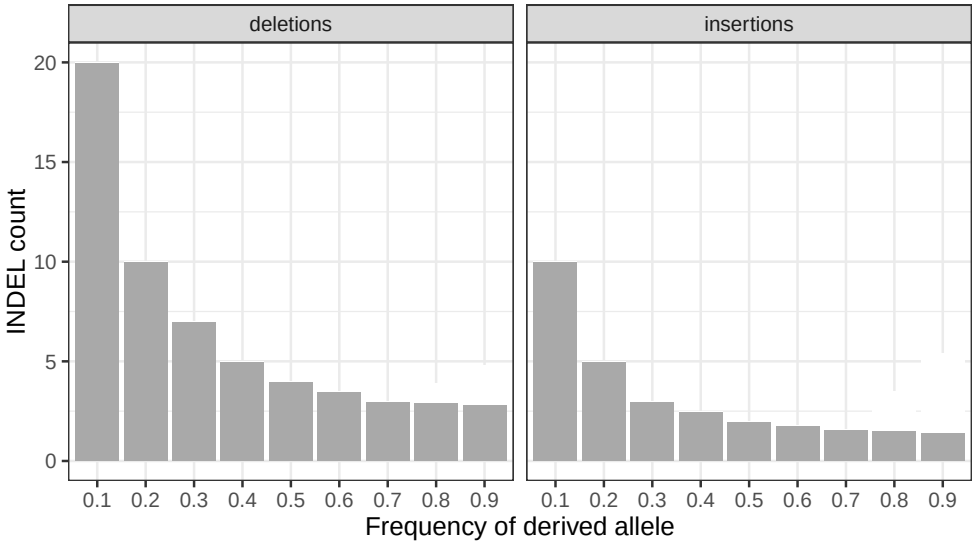
INDEL mutation

- Deletion bias in most organisms
- Polymerase slippage can explain majority of short INDEL events

INDEL selection

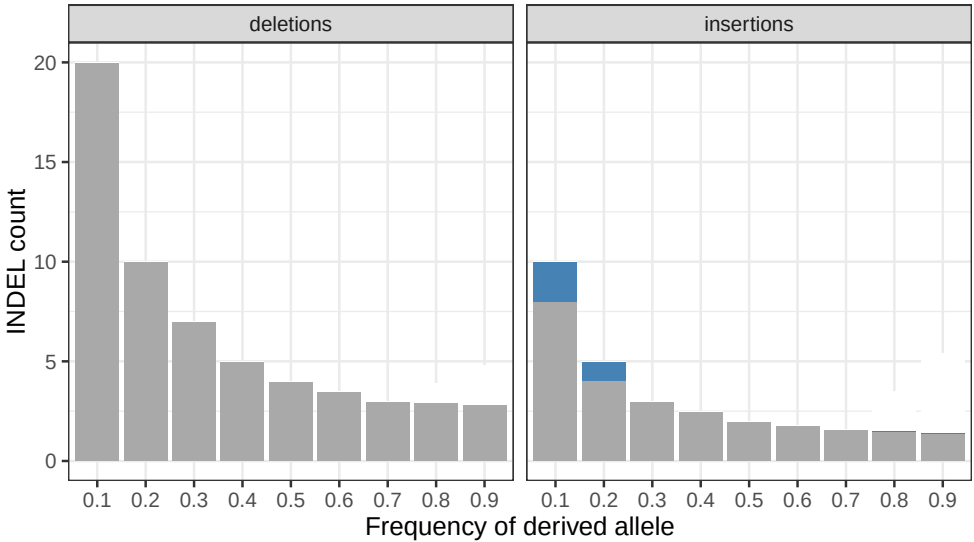
- Deletions
  - lower mean allele frequency
  - more deleterious
  - two breakpoints
- Insertions may be favoured:
  - elevated fixation probability
  - biased gene conversion
  - minimum intron size
  - polarisation error

Importance of polarisation error



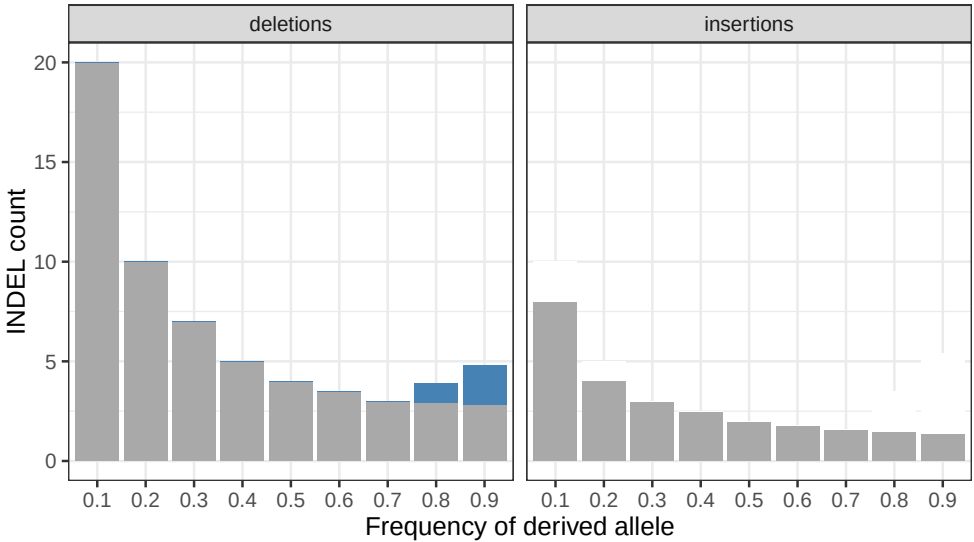
(see Hernandez et al., 2007)

Importance of polarisation error



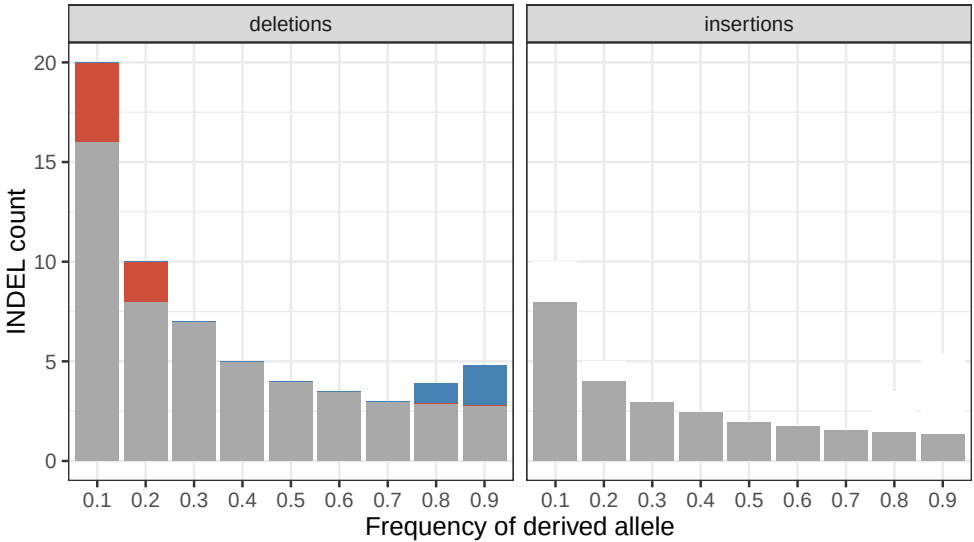
(see Hernandez et al., 2007)

Importance of polarisation error



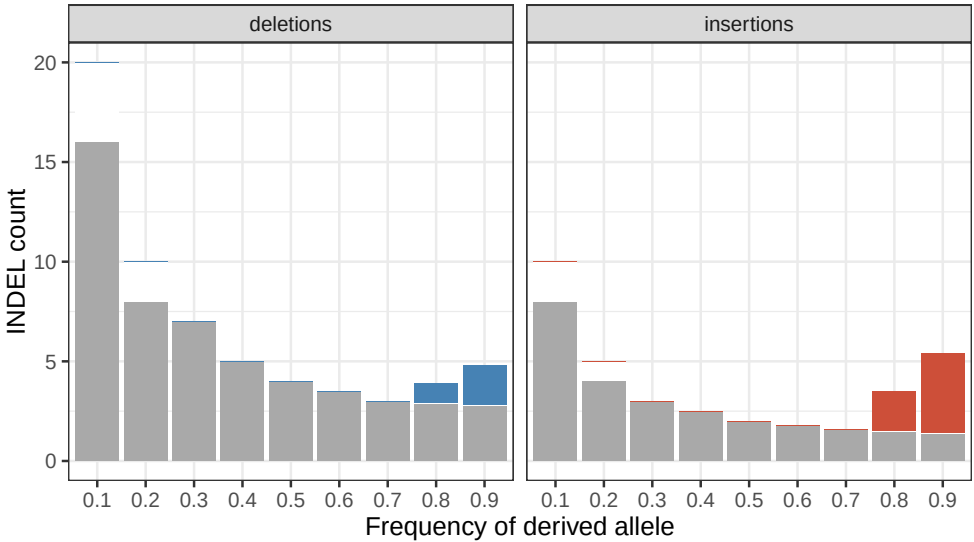
(see Hernandez et al., 2007)

Importance of polarisation error



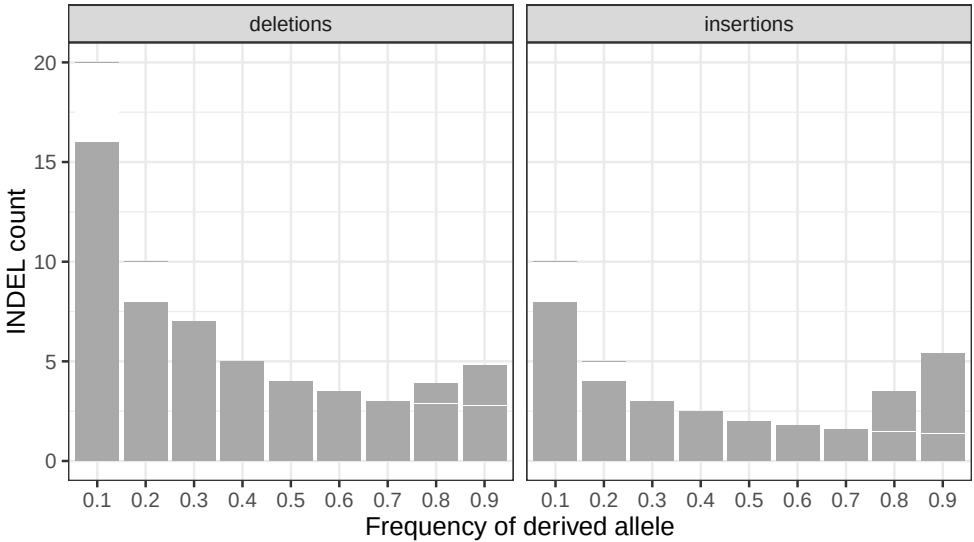
(see Hernandez et al., 2007)

Importance of polarisation error



(see Hernandez et al., 2007)

Importance of polarisation error



(see Hernandez et al., 2007)

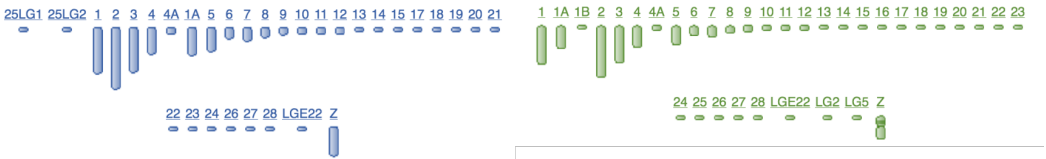
Aims

Quantify how natural selection shapes INDEL diversity in the great tit (*Parus major*)

1. within coding regions
2. in non-coding regions

Advantages of an avian system

- Conserved karyotype and synteny - good for alignments
- Genomes consist of few large macrochromosomes and many small microchromosomes
- Results in a highly dynamic recombination landscape - power to associations with recombination

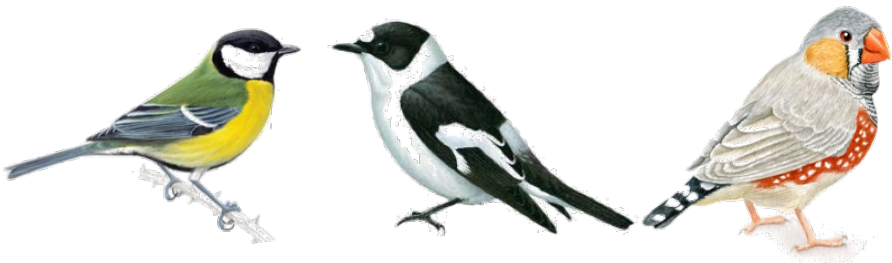


(van Oers et al., 2014; Stapley et al., 2008)

Data

Sample and pipeline

- 10 european great tit males (Corcoran et al., 2017)
- high coverage (44x)
- variant calling with GATK
- multispecies alignment between zebra finch, flycatcher and great tit
- parsimony based polarisation



The model - ‘anavar’

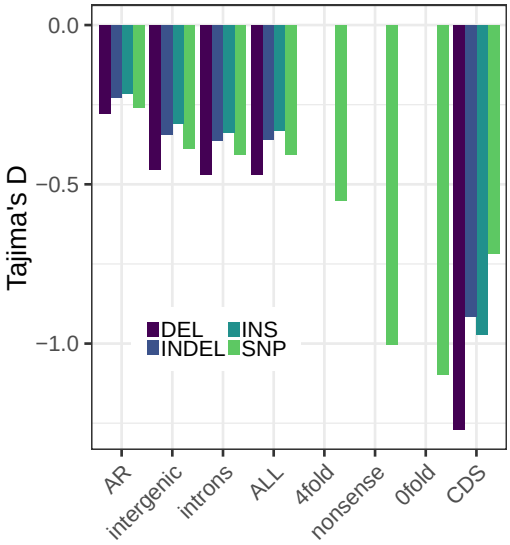
A novel maximum likelihood approach

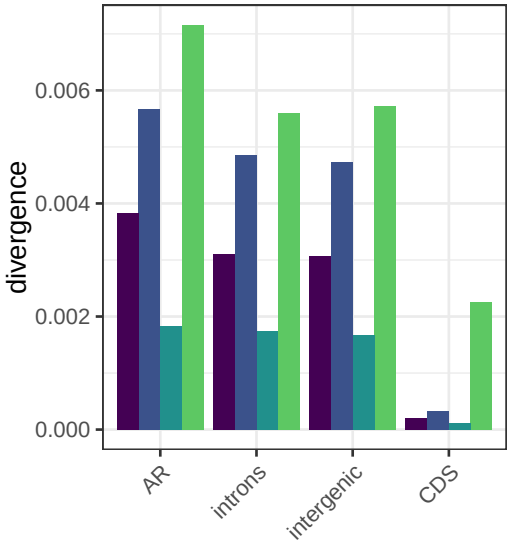
- takes the unfolded site frequency spectrum
- estimates for both insertions and deletions:
  - mutation rate ($\theta = 4N_e\mu$)
  - selection:
    - either selection coefficient ($\gamma = 4N_e s$)
    - or scale and shape paramater for distribution of fitness effects
  - polarisation error
- Controls for demography using neutral sites (Eyre-Walker et al., 2006)

(Barton and Zeng, 2018)

Dataset summary

Regional variation in purifying selection

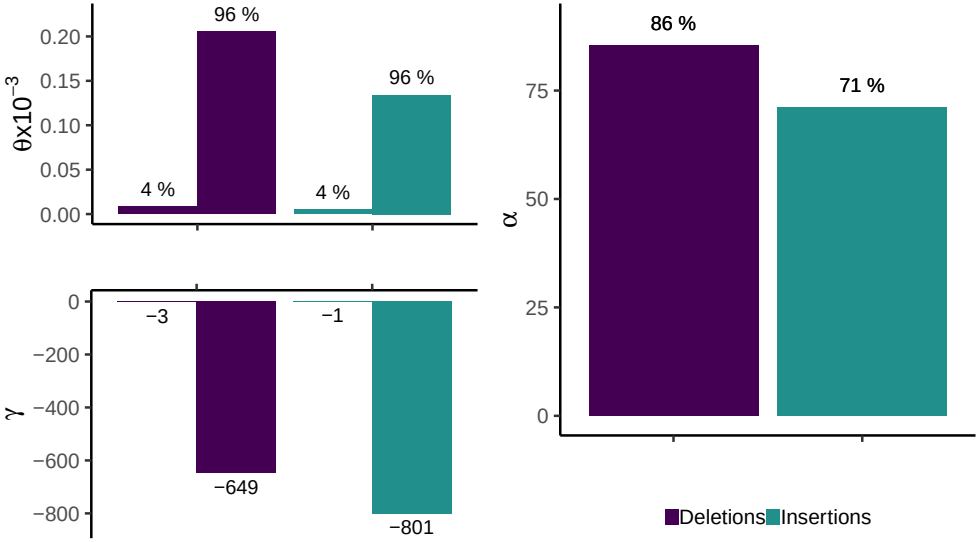




For SNPs: divergence = divergence estimate / 10

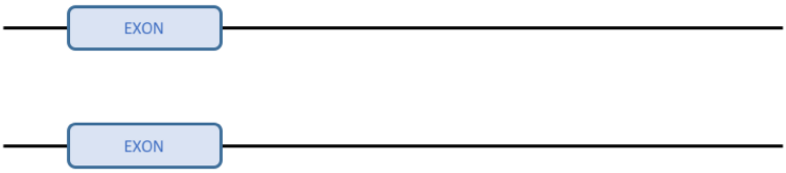
Coding INDELs

INDELs predominantly strongly deleterious

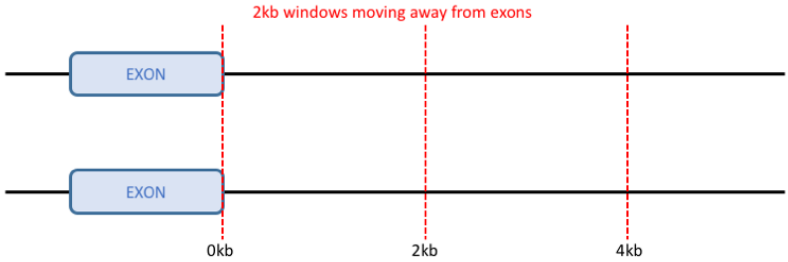


Moving away from coding regions

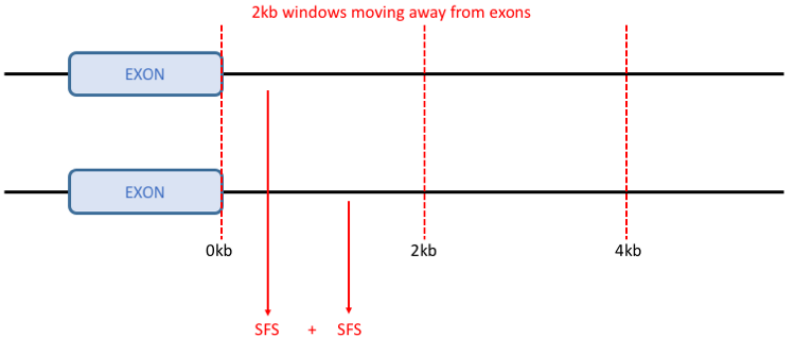
Approach



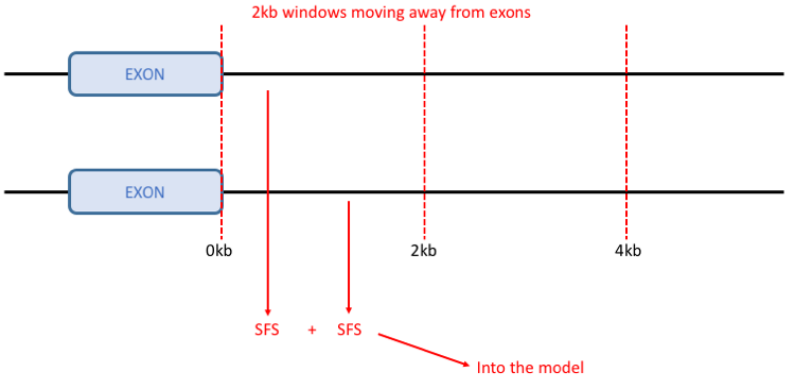
Approach



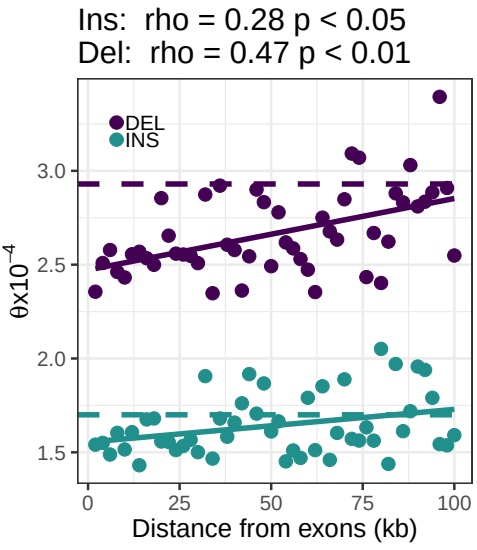
Approach

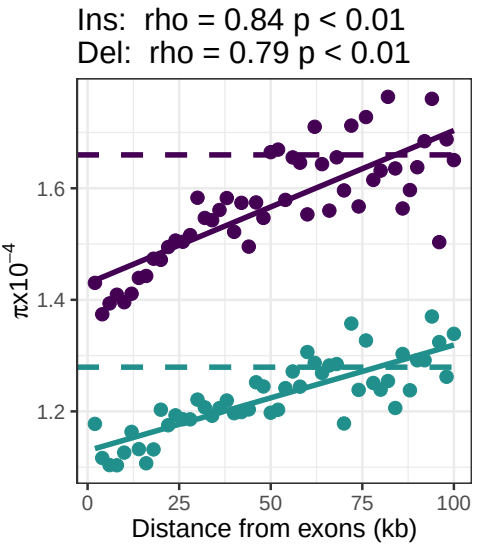

2kb windows moving away from exons
EXON
EXON
0kb
2kb
4kb
SFS
+
SFS

Approach


2kb windows moving away from exons
EXON
EXON
0kb
2kb
4kb
SFS
+
SFS
Into the model

Evidence for linked selection





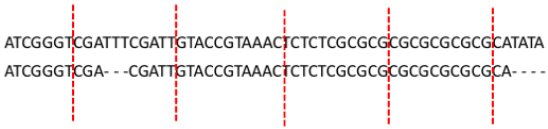
Recombination Analyses

Getting the data

ATCGGGTCGATTTCGATTGTACCGTAAACTCTCTCGCGCGCGCGCGCGCGCATATA

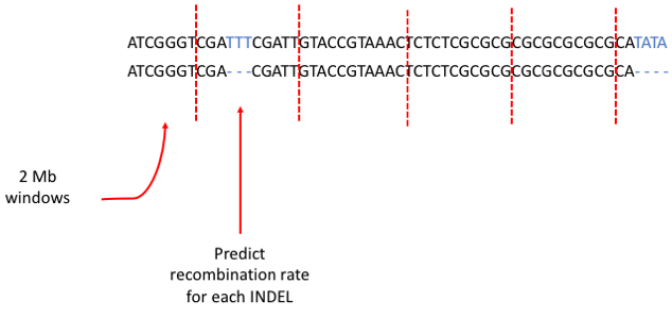
ATCGGGTCGA- - -CGATTGTACCGTAAACTCTCTCGCGCGCGCGCGCGCGCA- - - -

Getting the data

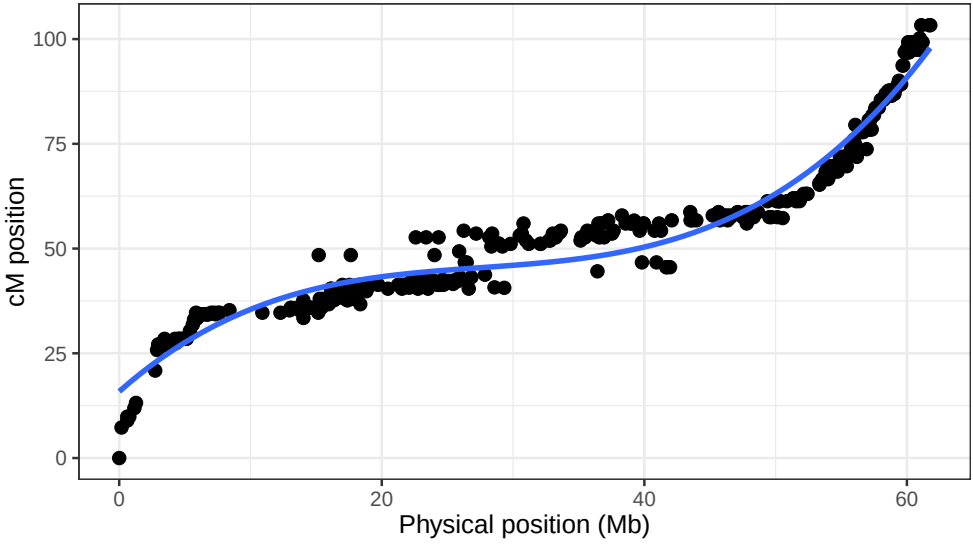


Getting the data

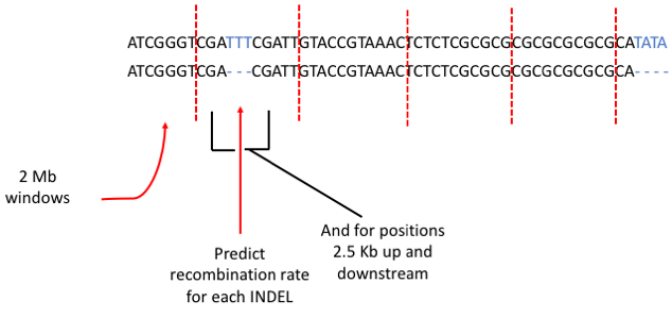
Getting the data



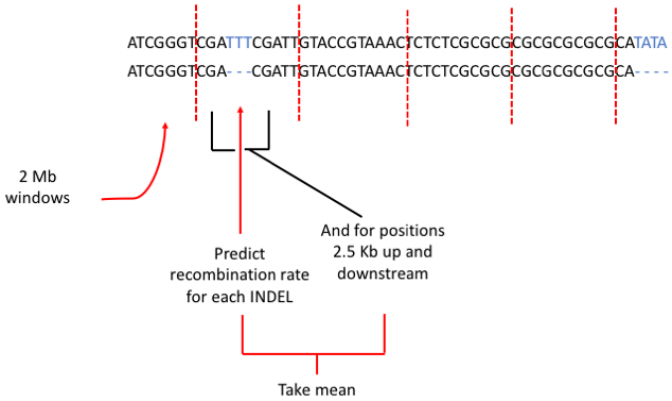
Getting the data - recombination rate



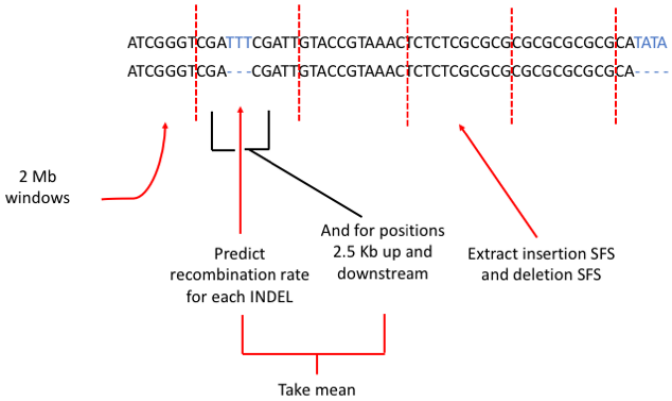
Getting the data



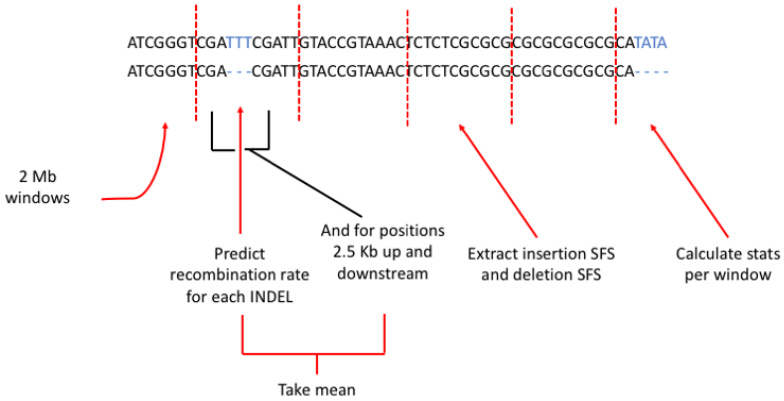
Getting the data



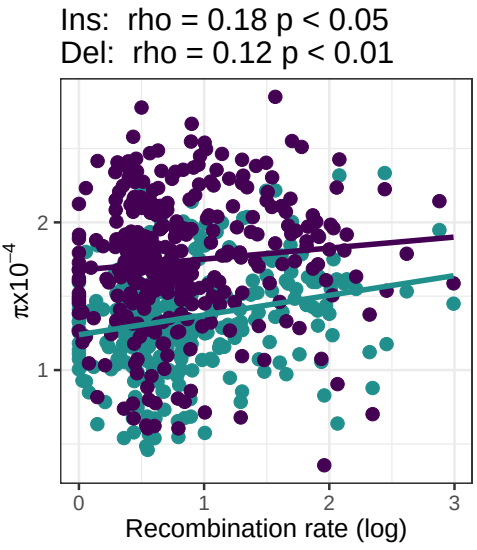
Getting the data

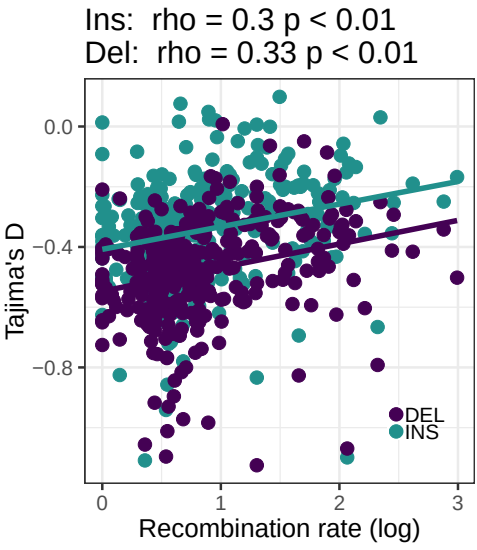


Getting the data



Association between diversity and recombination





Round up

Conclusion

- INDELs in genes mostly extremely deleterious - 96%
- Remainder are weakly deletrious - deletions more so
- α estimate at 71% and 86% for insertions and deletions
- Regions adjacent to exons have reduced INDEL diversity - genetic hitch-hiking
- Extends over relatively large distance 0-100kb
- Also areas of low recombination show reduced diversity

Next steps

- Intersting to investigate if reduced diversity is is due to postive selection or purrifying selection

Questions?